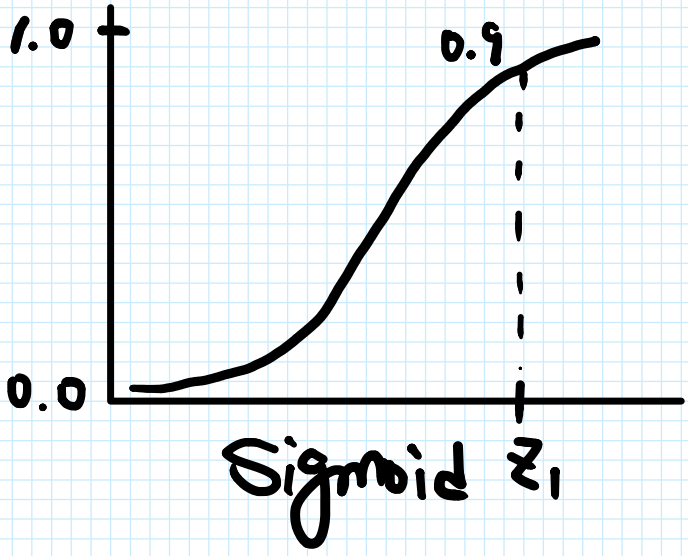
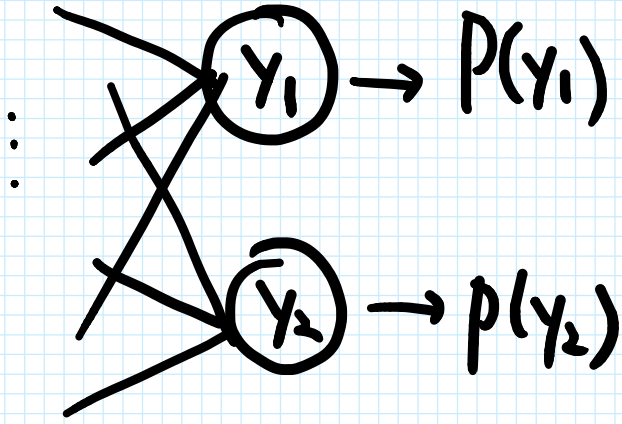
Day 3

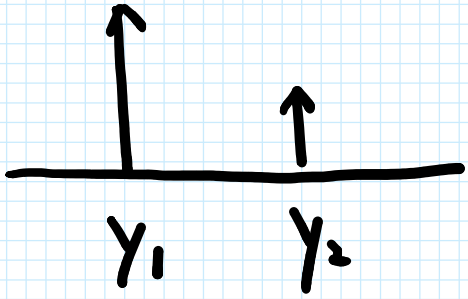
2018년 10월 17일 수요일 오전 9:08

분류 출력 처리



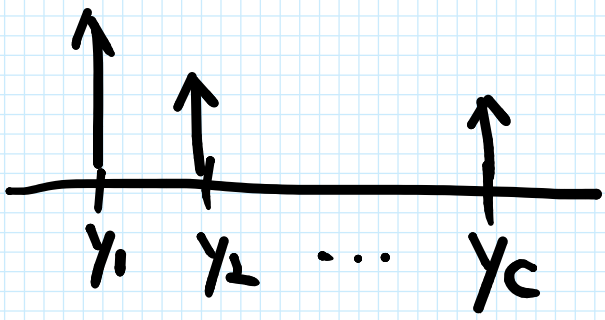


$$P(y_1) = 1 - P(y_2)$$



Bernoulli Dist.

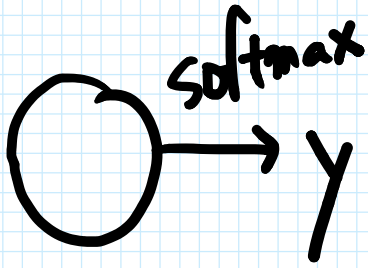
다중분류

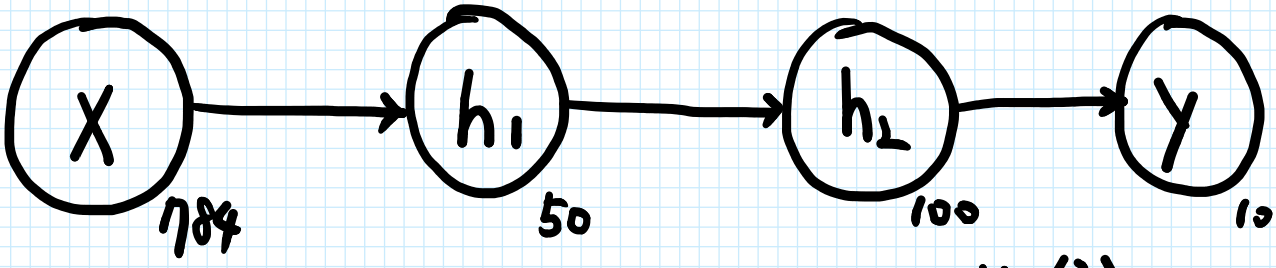


$$\sum_{k=1}^{c} P(y_k) = 1.0$$

$$0.0 \le P(y_k) \le 1.0$$

분류 분포

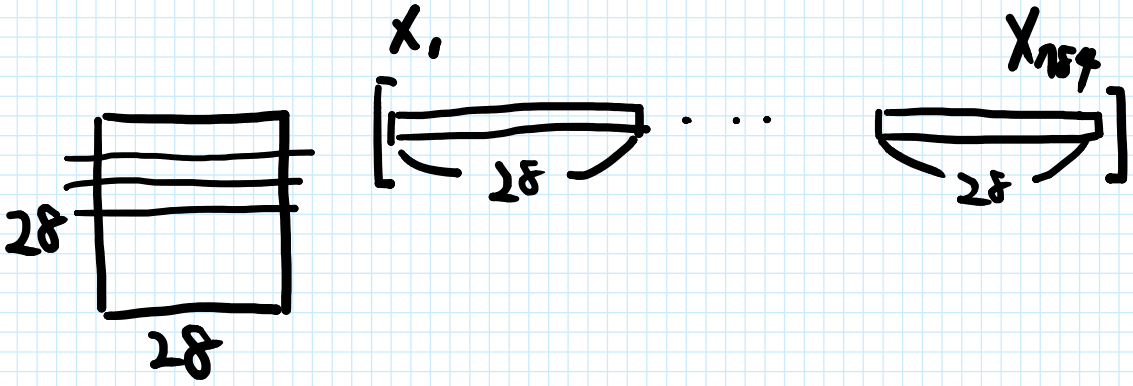




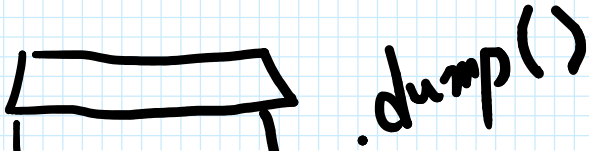
$W^{(1)}$ $(784, 50)$ $b^{(1)}$ $(50,)$

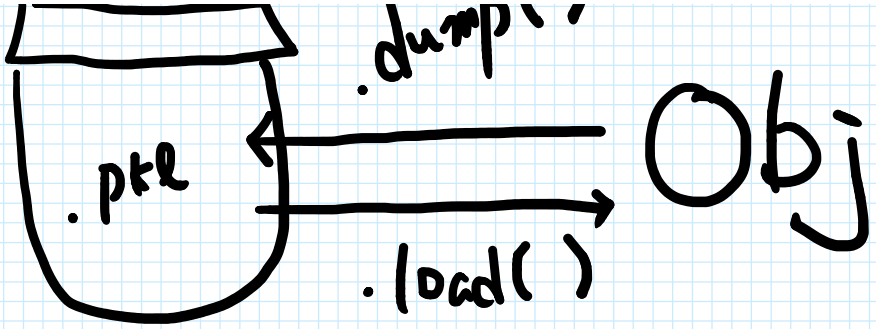
$W^{(2)}$ $(50, 100)$ $b^{(2)}$ $(100,)$

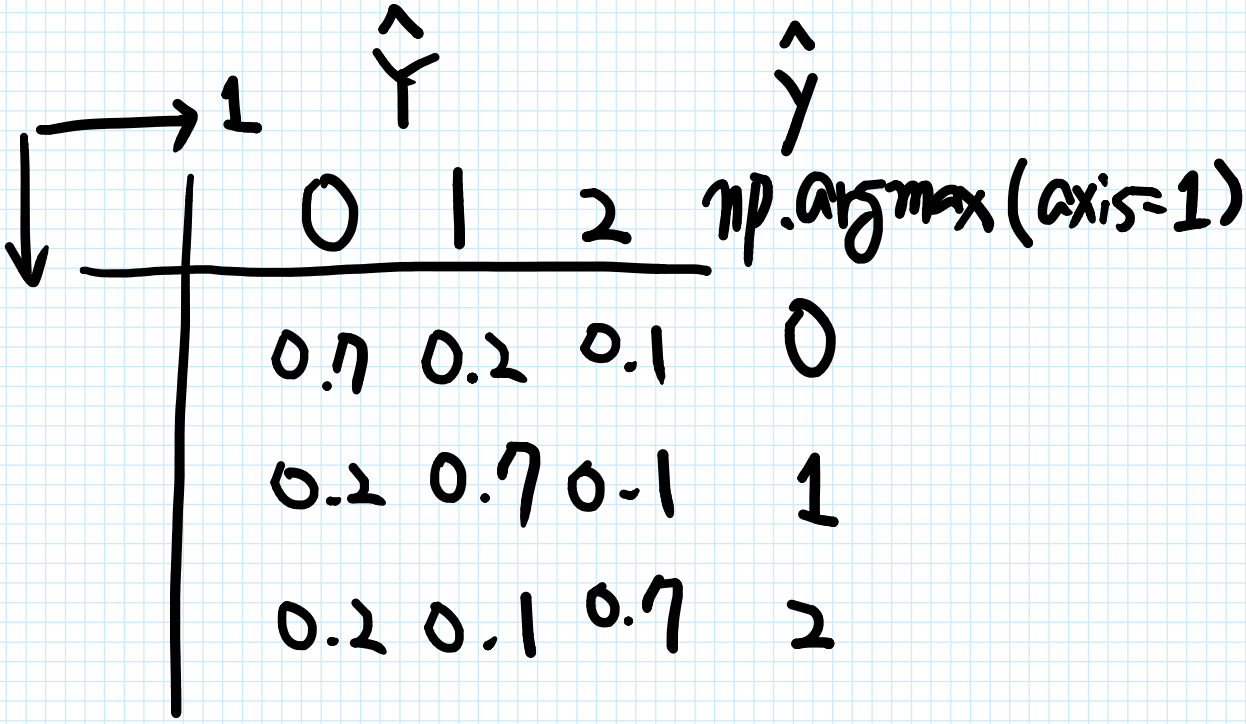
$W^{(3)}$ $(100, 10)$ $b^{(3)}$ $(10,)$



pickle







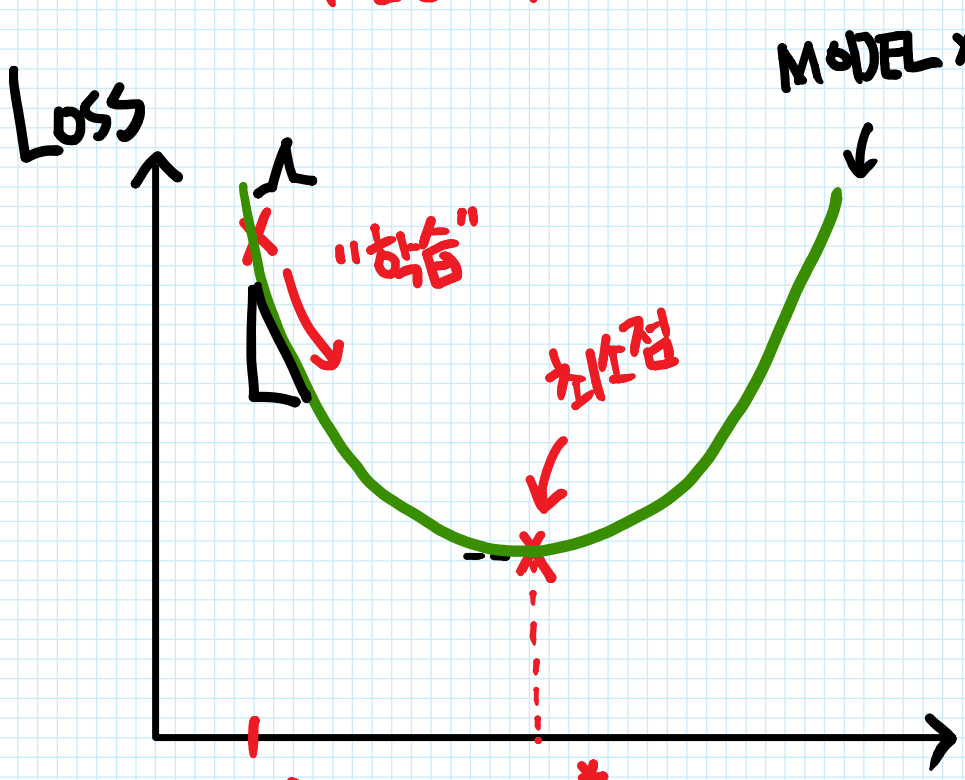
손실 함수 "종합오류"

✓ Loss = Cost = Objective function

회귀	분류
MSE	Cross Entropy Error

$$= \frac{1}{2} \sum_i (y_i - \hat{y}_i)^2 \qquad = -\sum_i y_i \log \hat{y}_i$$

각 샘플 오차

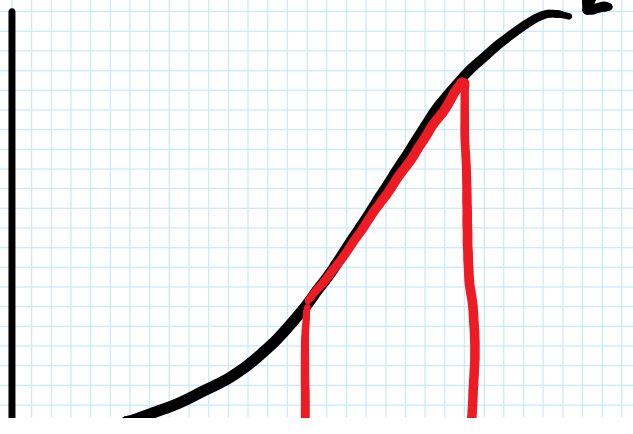


$Loss(\hat{y}, y)$

$\hat{y} \leftarrow$ MODEL W, b

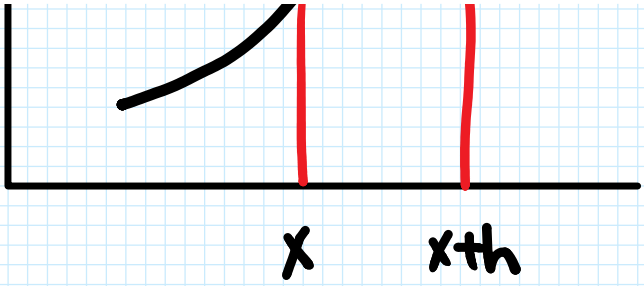
최적 → W^*

$$W^* = \underset{W}{\operatorname{argmin}}\ Loss(W)$$

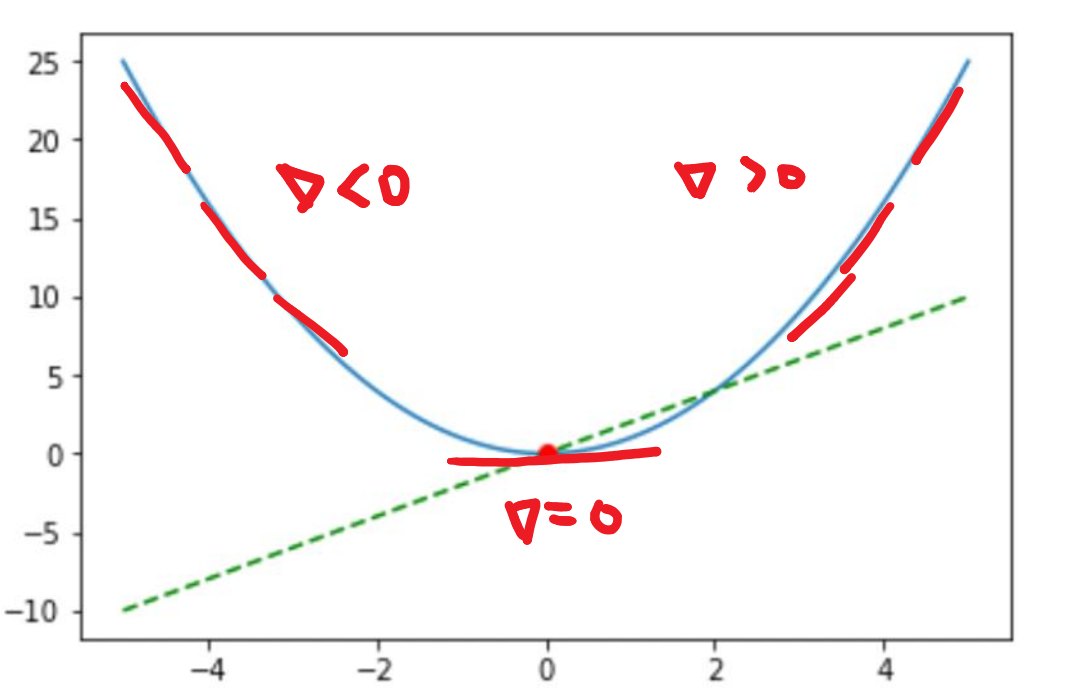


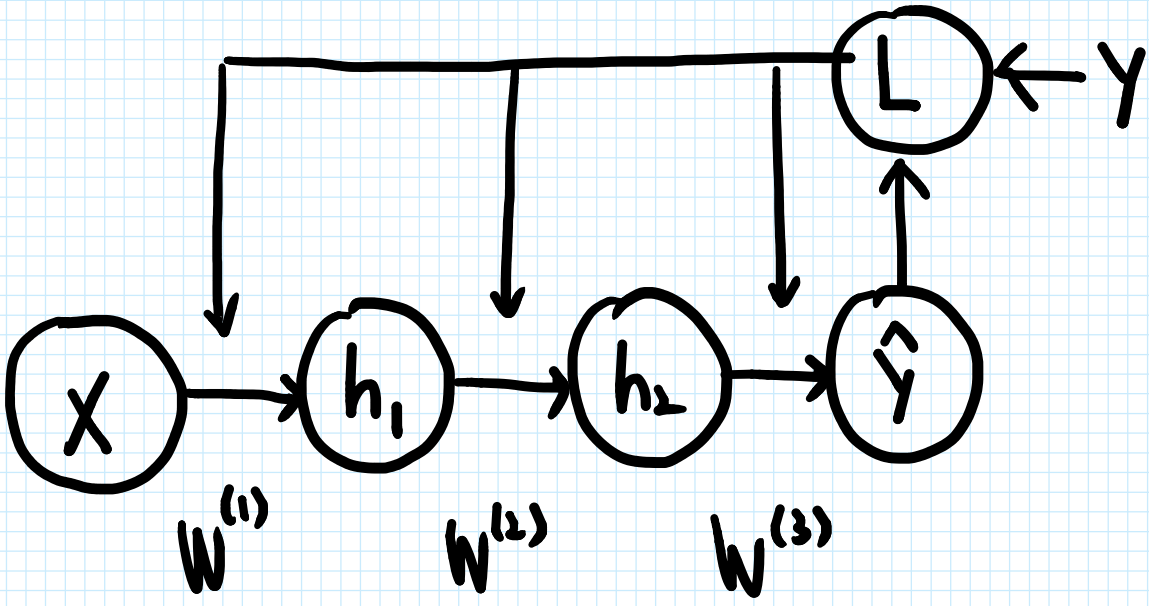
경사 $= \dfrac{f(x+h) - f(x)}{h}$

$h \to 0$ $\dfrac{f(x+h) - f(x)}{h}$



미분 $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$





Loss(y, $\hat{y}$)

$y = x^2$) 해석적

Loss (y, $\hat{y}$)

$\frac{dy}{dx} = 2x$ ↓ 해석적

$$\hat{y} = A(a_3 W^{(3)T} + b^{(3)})$$

$$= A((a_1 W^{(2)T} + b^{(1)} + b^{(3)})$$

$$\vdots$$

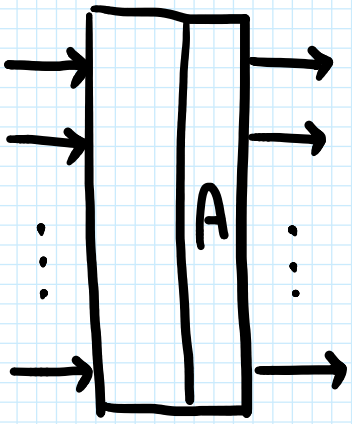
$\frac{dL}{dW}$ 손실 함수 변화 / 가중치의 변화

편미분 → 경사

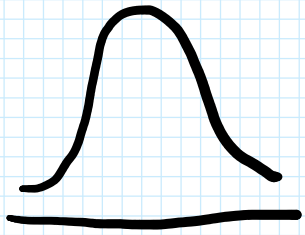
$$\begin{bmatrix} W_{11} & W_{21} \\ W_{12} & W_{22} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial W_{11}} & \frac{\partial L}{\partial W_{21}} \\ \frac{\partial L}{\partial W_{12}} & \frac{\partial L}{\partial W_{22}} \end{bmatrix}$$

은닉층



활성화값



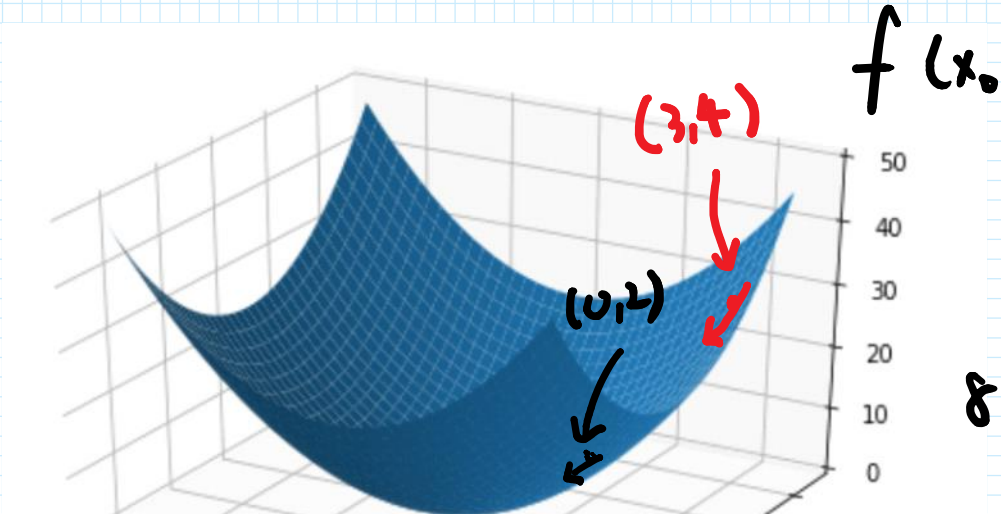
$$\hat{y} = A(z)$$

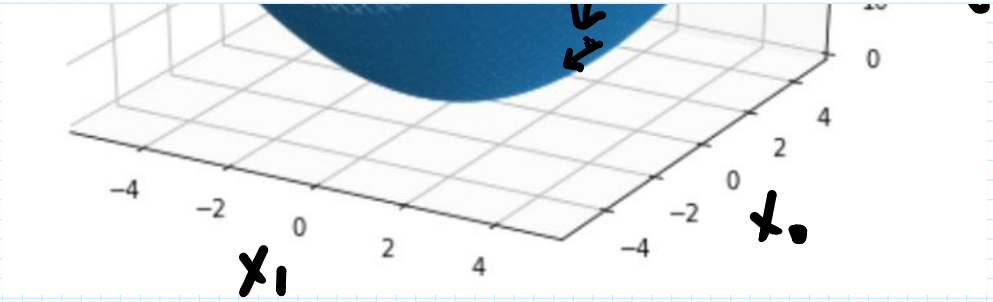
$$= A(A(z))$$

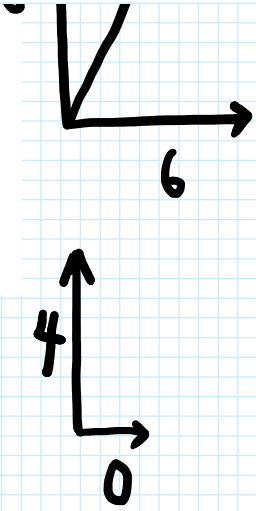
$$= A(A(A(z))) = z$$

1) 비선형적

$A(x) = x$







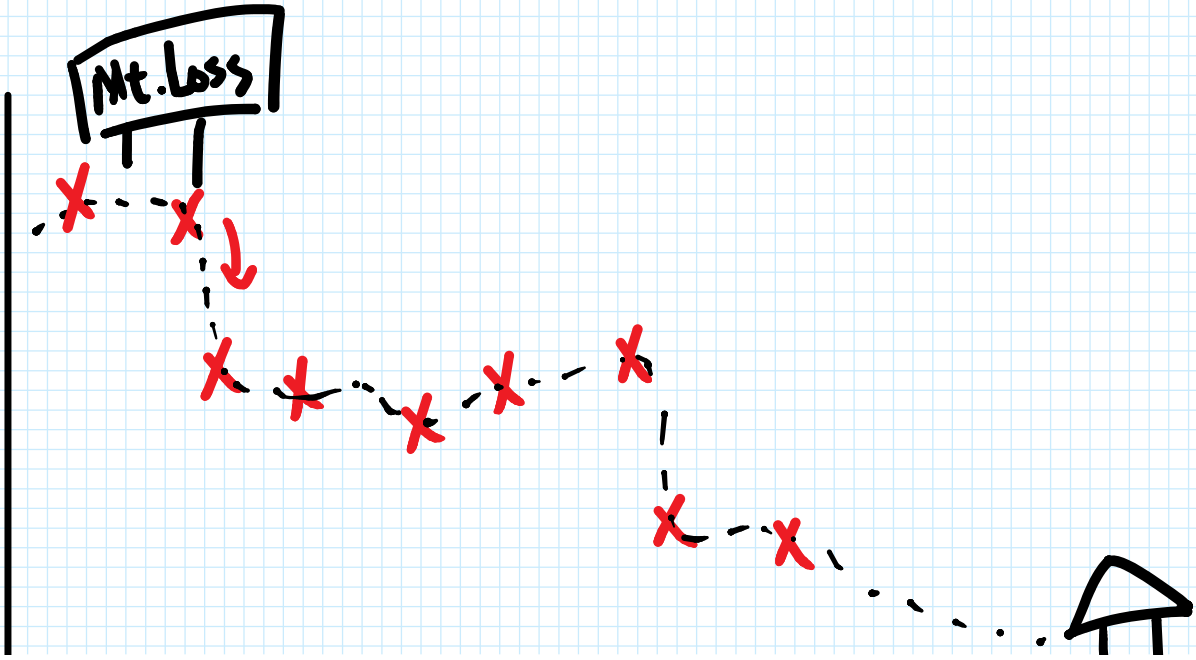
$$f(x_0, x_1) = x_0^2 + x_1^2$$

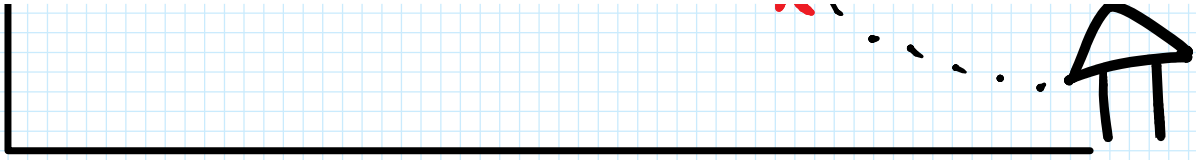
$$\frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial x_0}, \frac{\partial f}{\partial x_1} \right) = \nabla f$$

경사(Gradient)

경사하강법

Gradient Descent (GD)

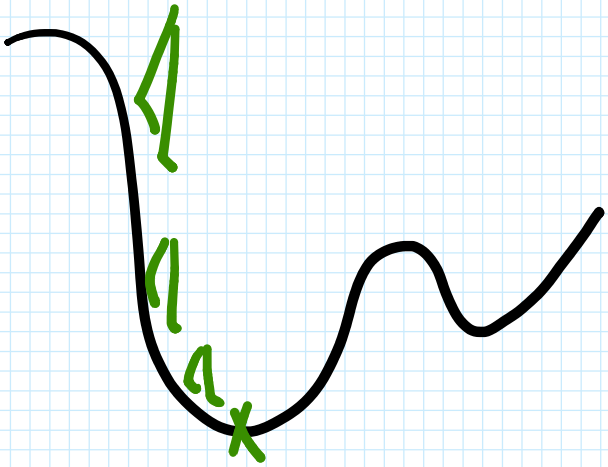




$$x' \leftarrow x - \Lambda \nabla f$$

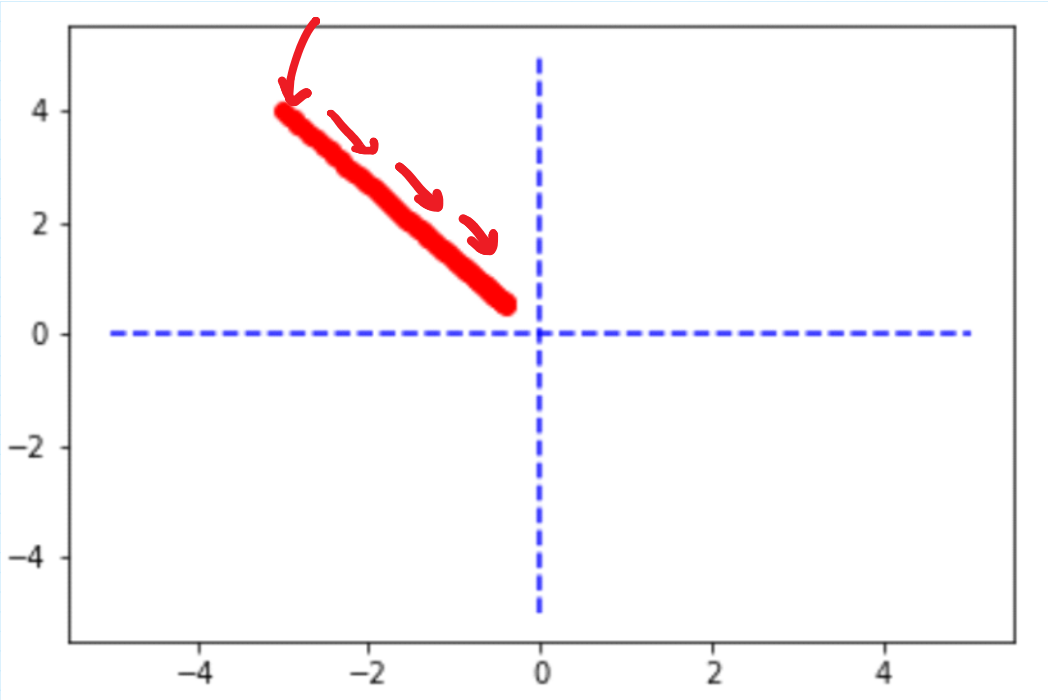
경사

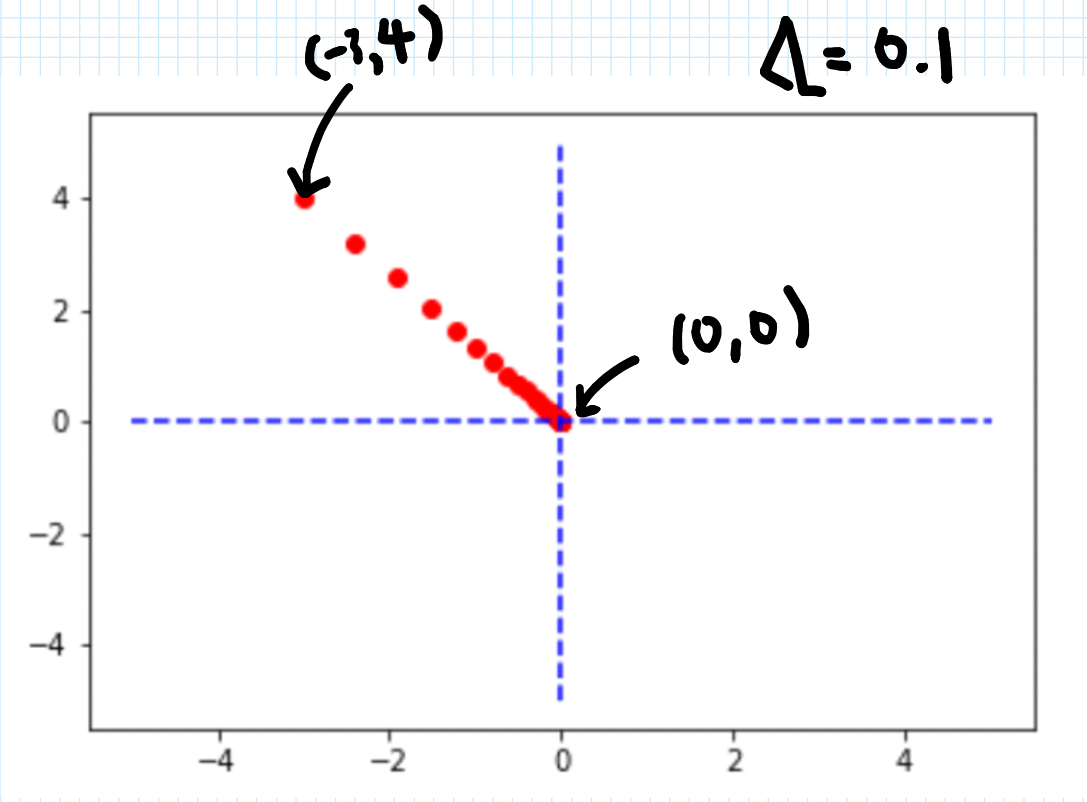
학습률



(-3, 4)

$\Lambda = 0.01$





$$Loss(\hat{y}, y)$$

Gradient (f, X)

np.argmax()

사진 ID	Y	y_0	y_1	y_2	...	y_9	label

One Hot Encoding (OHE)

np.argmax( )

	R?	G?	B?
R	1	0	0
G	0	1	0
B	0	0	1

R < G < B

1986 오차역전파

$$\text{가격} = \left\{(\text{사과} \overset{+1}{\nearrow} \times \text{개수}) + (\text{귤} \times \text{개수})\right\} \times \text{부가세}$$

$$715 = \left\{(100 \times 2) + (150 \times 3)\right\} \times 1.1$$

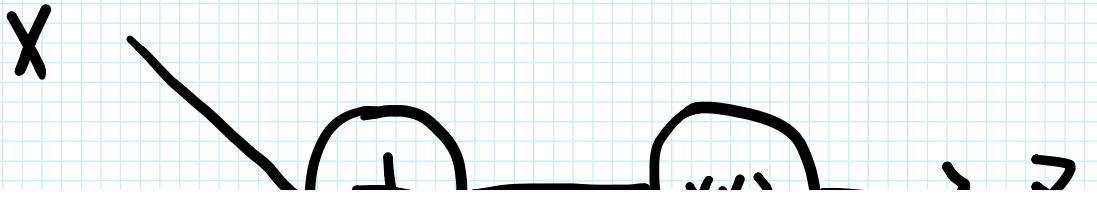
$$\frac{\partial L}{\partial \text{사과}} = 100 \times 2 + 150 \times 3$$

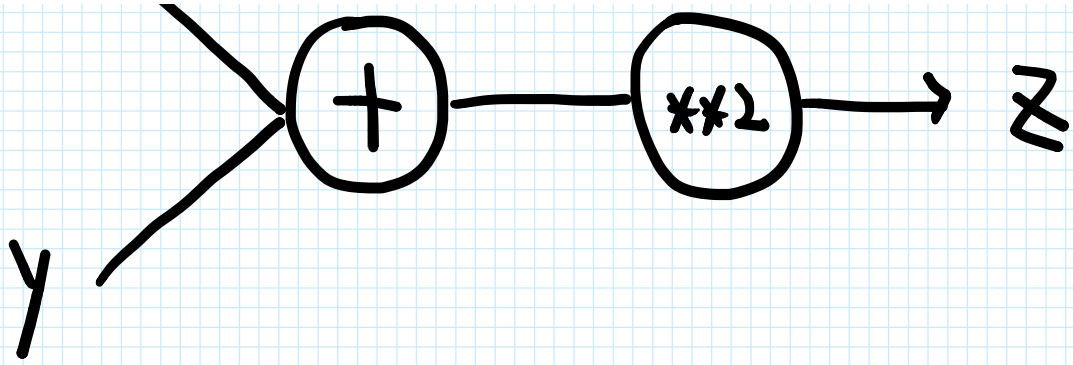
연쇄법칙

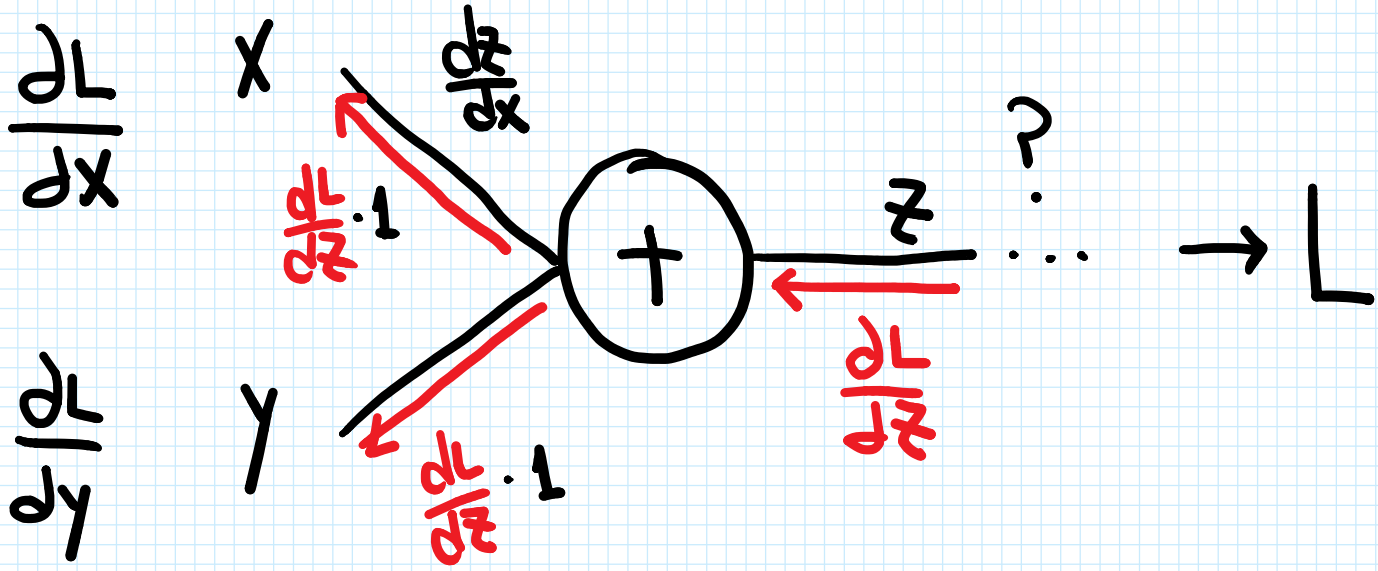
$$z = (x+y)^2 \quad \frac{\partial z}{\partial x} \quad \frac{\partial z}{\partial y}$$

$$z = t^2$$

$$t = x+y$$







$$z = x + y$$

$$\frac{\partial z}{\partial x} = 1 \quad , \quad \frac{\partial z}{\partial y} = 1$$

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial x} = \frac{\partial L}{\partial z} \cdot 1$$

$$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial y} = \frac{\partial L}{\partial z} \cdot 1$$

$$\frac{dL}{dy} = \frac{dL}{dz}\frac{dz}{dy} = \frac{dL}{dz} \cdot 1$$